



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of : Mahmoud K. Jibbe
Serial No. : 10/629,877
Filed : July 29, 2003
Art Unit : 2188
Examiner : Doan, Duc T.
Title : A METHOD FOR ESTABLISHING A REDUNDANT
ARRAY CONTROLLER MODULE IN A STORAGE ARRAY
NETWORK

**DECLARATION OF PRIOR INVENTION IN THE UNITED STATES OR IN A
NAFTA OR WTO MEMBER COUNTRY TO OVERCOME CITED PATENT OR
PUBLICATION (37 C.F.R. § 1.131)**

Dear Sir:

This declaration is to establish conception of the invention at a date prior to July 23, 2003, which is the earliest effective date of a reference (U.S. Patent Publication Number 20040139365) cited by the examiner, coupled with diligence in completing the invention in this application in the United States until filing of the patent application on July 29, 2003.

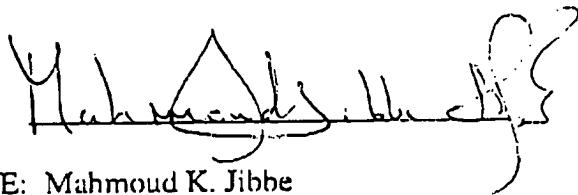
1. The person making this declaration is the inventor.
2. Conception of the invention occurred prior to the effective date of the reference, July 23, 2003.
3. From the date of conception, the inventor diligently prepared an invention disclosure (Exhibit A).
4. The invention disclosure, prepared by the named inventor was received by the Intellectual Property Department of LSI Logic Corporation as shown on the first page of the invention disclosure (Exhibit A).
5. The invention disclosure was sent to Outside Counsel for draft of the application.
6. The patent application was filed on July 29, 2003.

DECLARATION

As a person signing below:

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Dated: 12/21/2005



FULL NAME: Mahmoud K. Jibhe

EXHIBIT A

LSI LOGIC CORPORATION INVENTION DISCLOSURE FORM
(Mail to "Intellectual Property Law Dept. Attn: New Invention Disclosure" – Mail Stop AD-106)
Please attach additional pages where required

1. Title of the Invention: A method for establishing Redundant Array Controller Module in a Storage Array Network	Number of Attached Pages: 5
If invention was presented in LSI "Brainstorming Session", please specify docket number assigned to you in reminder email -	
2. List of Inventors: (Please provide following - your name, email, phone and fax numbers) Mahmoud K. Jibbe; <u>mahmoud.jibbe@lsil.com</u> (316) 636-8810; (316) 636-8321	
3. Public Disclosure of the Invention: <ul style="list-style-type: none">▪ Was invention ever disclosed, either orally or in writing? If yes, please provide details, such as when and to whom. No▪ Are there plans to disclose the invention in the future? If yes, please provide details, such as when and to whom. No	
4. Use of the Invention: <ul style="list-style-type: none">▪ Has invention been used? If yes, please provide details, such as when and how. No▪ Are there any plans to use the invention in the future? If yes, please provide details, such as when and how. No immediate plans for inclusion in a product. Future inclusion is possible.	
5. Invention "offered for" or "on sale": <ul style="list-style-type: none">▪ Was a product/process containing invention "offered for sale" or "sold?" If yes, pl. provide details, such as when & to whom. No	
6. Date of your Invention: <ul style="list-style-type: none">▪ Specify the earliest date when you conceived of your invention: (e.g. the conception date)▪ Specify the earliest date a prototype of the invention was built: (e.g. reduction to practice)	
7. Background of the Invention: (please attach additional pages) <ul style="list-style-type: none">▪ The field to which your invention pertains. See attached pages.▪ The problems that your invention aims to solve. See attached pages.▪ Existing approaches toward solving those problems (if any). See attached pages.▪ What are these existing approaches lacking? See attached pages.	
8. Detailed Description of Your Invention: (please attach additional pages) <ul style="list-style-type: none">▪ Provide enough information and detail so that another person in your field could make and use your invention.▪ If available, provide description with any existing reports, presentations, e-mails, sketches, drawings, schematics, photos, etc.▪ At least one simple Figure or Flowchart of your invention MUST be included.▪ Identify the new features of your invention.▪ List the advantages of your invention.▪ Disclose alternate ways of making and using of your invention.	
9. Prior Art: List only those patents, products, processes, journal articles, presentations, conferences, seminars, and other knowledge <u>that you are aware of</u> (you have no duty to conduct a search), that are related to the subject matter of your invention: See attached pages.	

LSI LOGIC CORPORATION INVENTION DISCLOSURE FORM
(Mail to "Intellectual Property Law Dept. Attn: New Invention Disclosure" – Mail Stop AD-106)
Please attach additional pages where required

10. Customer/Vendor Contracts:

Was invention developed during performance of a customer/vendor contract? If yes, please provide details. No

11. Government March-In Rights:

Was invention conceived during performance of government contract? If yes, please provide details. No

12. Signatures:

I/We the aforementioned inventor(s) of this invention, which was developed during the course of employment, submit in confidence this invention disclosure to attorneys of the LSI Logic Legal Dept. for the purpose of obtaining a legal opinion and/or advice as to availability of patent, trade secret, and/or copyright protection related to the material contained within.

I(We) believe myself(ourselves) to be the first and original inventor(s) of this invention:

Inventor (s)

Mahmoud Jibbe

Date:

Two Witnesses who have read and understood this invention disclosure:

GROVER P. HILLIPS *Grover P. Phillips*
Ken Fugate *Ken Fugate*

Date:

A method for establishing Redundant Array Controller Module in a Storage Array Network

Inventor: M. K. Jibbe

Date: [REDACTED]

Abstract:

The method of this invention is to establish redundancy at the array controller module level. The redundancy is established via a state machine, which locks step the array controller modules and a switch in a quick loop topology (Internal or external to the drive trays). The state machine provides the redundant array controller module multiple paths to the drive trays. These paths allow the redundant array controller the capability of taking over the I/O configuration and I/O access to the drive trays in the case of an array controller module failure (i.e. failover to the redundant array controller module).

Introduction:

In a Storage Array Network (SAN), an array controller module consists of two redundant array controllers. The two array controllers can be in a dual active mode, or an active – standby mode. In both modes, if the active controller fails then the other active array controller or the standby array controller will take over the failed active array controller responsibilities and activities. This is typically what the array technology referred to as redundancy. The current array technology does not handle the following cases:

- 1) Controller module failure, which prevents both array controllers from accessing the drive trays. These failures can be related to power supplies failure, Fans failure, mid plane failures, and others.
- 2) Active controller fails in such a way, which prevents the alternate controller from accessing the drive trays.
- 3) Standby controller fails in such a way, which prevents the active array controller from accessing the drive trays.

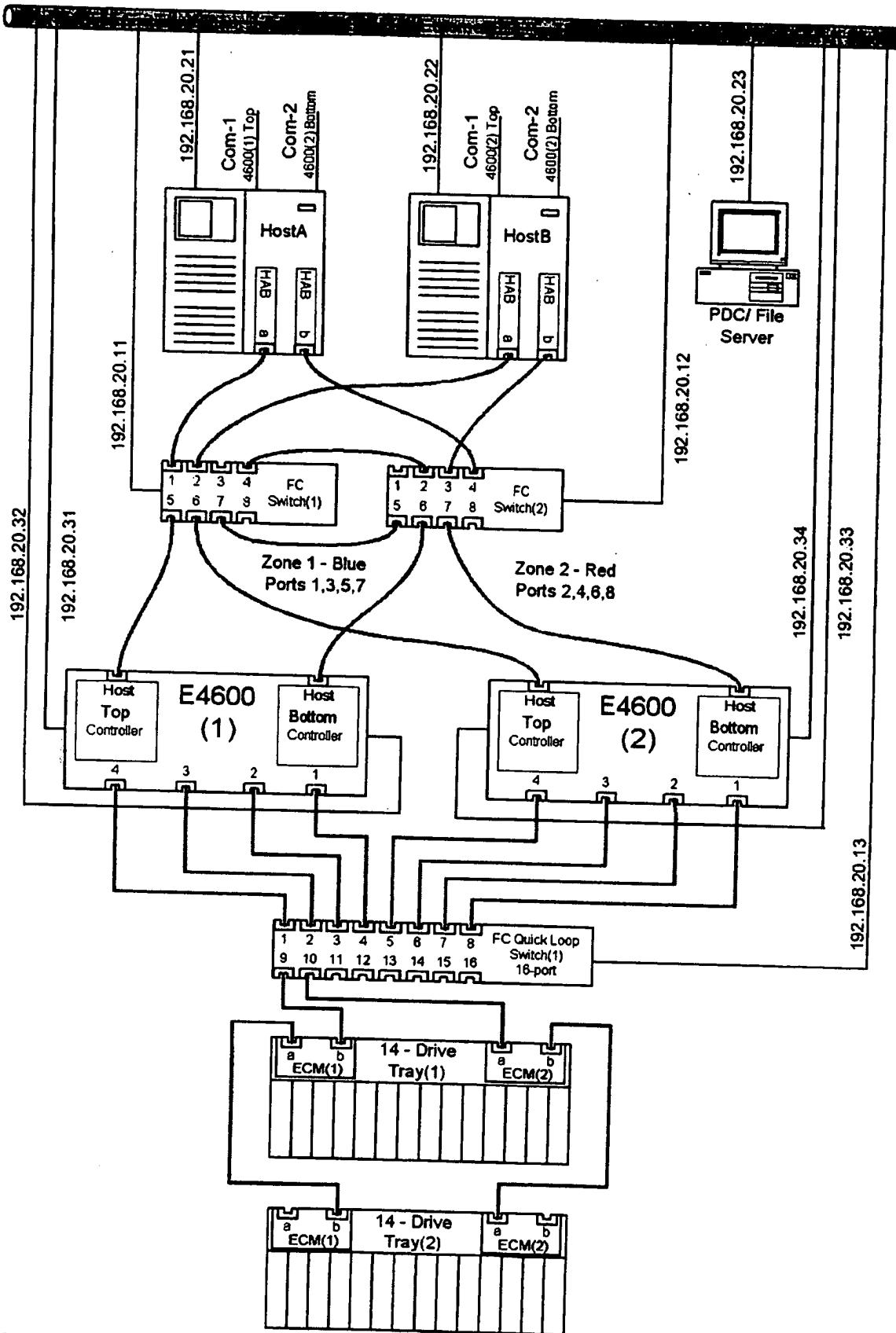
The method of this invention will be able to resolve the above problems by isolating the failed array controller module and transfer the responsibilities and activities to the alternate array controller module without changing the hardware design of the existing array controller module but with the addition of external / internal switch and a state machine for the lock step of the redundant alternate array controller modules.

Description of the Invention Method:

The method of this invention requires a switch (internal or external to the drive tray) between the array controller modules and the bank of the drive trays as shown in Figure 1.

Figure 1: System configuration for the method of this invention.

LSI LOGIC CORPORATION INVENTION DISCLOSURE FORM
 (Mail to "Intellectual Property Law Dept. Attn: New Invention Disclosure" – Mail Stop AD-106)
Please attach additional pages where required



Referring to Figure 1, the array controller module (1) mini hub 1 is connected to port 4 of the added switch. Port 9 of the added switch is connected to port (b) of the ESM (1) of tray (1). Port (a) of ESM (1) of tray (1) is connected to Port (b) of ESM (1) of tray (2). Port (a) of ESM (2) of tray (2) is connected to port (b) of ESM (2) of tray 1. Port (a) of ESM (2) is connected to port 10 of the added switch. Port 8 of the added switch is connected to the mini hub 1 of the array controller module (2). These connections establish one loop with

LSI LOGIC CORPORATION INVENTION DISCLOSURE FORM
(Mail to "Intellectual Property Law Dept. Attn: New Invention Disclosure" – Mail Stop AD-106)
Please attach additional pages where required

two redundant array controller modules access to the bank of the drive trays. The other mini hub connections provide the method of this invention redundant paths / loops to the bank of drive trays for two or more redundant array controller modules.

The method of this invention specifies that array controller A of module (1) and array controller A of module (2) belongs to the same multicast group (1), and array controller B of module (1) and array controller B of module (2) belongs to multicast group (2). The multicast groups are required by the method of this invention because host A and host B broadcast any command (I/O "exchange data" or symbol command "array system configuration") to the A controllers in group (1) and B controllers in group 2. This way the frames for the A (B) controller of Module 1 are forwarded to A (B) controller of module 2. This way if module (1) fails then module (2) will take over.

It should be noted that the symbol command to configure the array system could be sent from a client over the network or an agent over the fibre or SCSI cable. In both media / network, the concept of broadcasting the command to both array controller provides the redundant paths.

The method of this invention handle the failover as depicted in the following steps:

1) Case 1: Both array controller modules are healthy (Active – Standby mode)

In this case, upon receiving a command, the primary array controller module (PM) sends an ECHO frame to the secondary array controller module (SM) informing it that the command with exchange ID is processed. The secondary module sends an ACCEPT frame indicating the command with Exchange ID is dumped. In the mean time, the secondary controller periodically sends an ECHO frame to check the health of the primary module. The primary module sends an ACCEPT frame to indicate good status. With this method, the primary and secondary are locked step in term of processing commands. It should be noted since the secondary module, primary module, and bank of drive trays are on the same redundant loops, the secondary module monitors all the configurations and I/O access of the primary module (monitor state).

2) Case 2: Both array controller modules are healthy (Active – Active mode)

In this case, the hosts broadcast a command to both array controller modules. The array controller module, which is capable for processing the command, sends an ECHO frame to the other array controller module to inform it that the command with an exchange ID is processed to the drive trays such as cache flushing. The alternate array controller module sends an ACCEPT frame to drop the command with exchange from its queue. In this mode, the volume owner to be accessed gets transferred to the available array controller module without any user intervention and as a result the overall performance of the system is almost doubled.

3) Case 3: Primary Array Module fails and Secondary Array Module takes over

In this case, the secondary module sends an ECHO frame to the primary controller but did not an accept frame after Error Detection Timeout expires. The secondary module will disable the primary module switch ports, namely ports 1, 2, 3, and 4. This way regardless how the primary module fails, namely cases 1, 2, and 3 presented previously, the secondary module can access the bank of drive trays. Then, the secondary module completes any outstanding exchanges.

4) Case 4: Primary Array Module recovers and Secondary Array Module fails back.

In this case, the primary controller comes online and sends ECHO frames to controller A and B of secondary indicating online state using the switch (or loop) between hosts and array modules. The secondary module processes any outstanding exchanges and returning ACCEPT frames to the A and B controllers of the primary module indicating a fail back state. Upon receiving the ACCEPT frames, the A and B controllers of the primary module assume the responsibility of processing exchanges.

LSI LOGIC CORPORATION INVENTION DISCLOSURE FORM
 (Mail to "Intellectual Property Law Dept. Attn: New Invention Disclosure" – Mail Stop AD-106)
Please attach additional pages where required

For additional information, refer to the flow chart in Figure 2

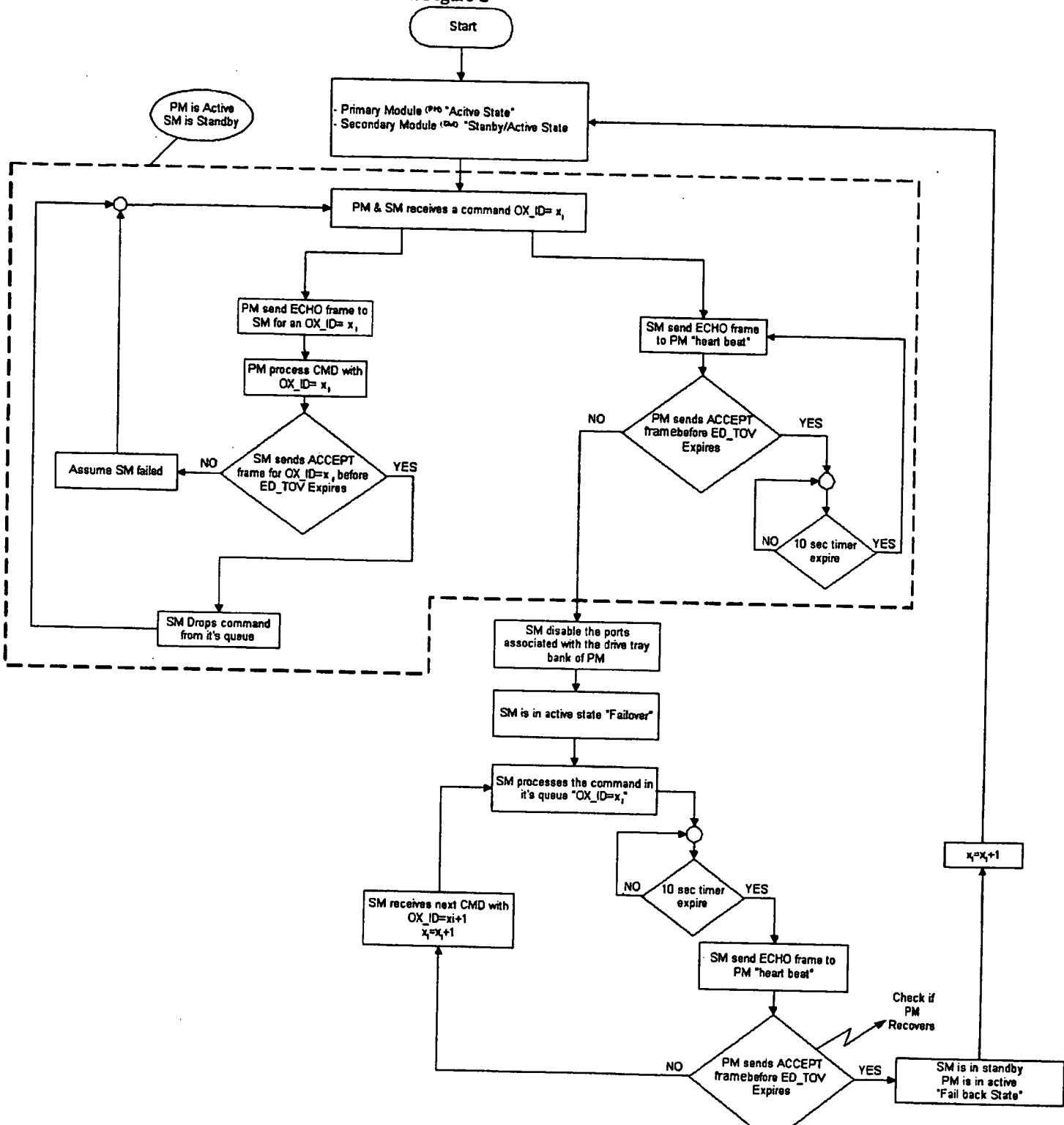


Figure 2: Basic state machine for the method of this invention.

LSI LOGIC CORPORATION INVENTION DISCLOSURE FORM
(Mail to "Intellectual Property Law Dept. Attn: New Invention Disclosure" – Mail Stop AD-106)
Please attach additional pages where required

Rules and States of this method:

The method of this invention adopts the following rules / States for each mode of operation:

I. Active – Standby Mode:

- 1) PM (SM) sends an ECHO frame to the SM (PM) is equivalent to Controller A of PM (SM) sends an ECHO frame to Controller A of SM (PM), and Controller B of PM (SM) sends an ECHO frame to Controller B of SM (PM).
- 2) In the optimal state, the PM controller processes the commands and the SM is in the standby state. In this state the controller A and B of PM return ECHO frames to controller A and B of SM with ECHO data set to healthy state.
- 3) In the Standby state, the SM sends an ECHO frame to the PM and get an ACCEPT frame before the Error Detection Time Out Value (ED-TOV), which is 2 seconds, expires.
- 4) PM (SM) is failed or in the failed state if controllers A and B of PM (SM) did not return an ACCEPT frame to the SM (PM) with the ECHO data set to healthy state.
- 5) PM (SM) is degraded or in the degraded state if controllers A or B of PM (SM) did not return an ACCEPT frame to the SM (PM) with the ECHO data set to healthy state.
- 6) In the fail Back state, the SM sends an ECHO frame to the PM and get an ACCEPT frame before the Error Detection Time Out Value (ED-TOV), which is 2 seconds, expires. In this transition state, the PM transfer to the optimal or degraded state and starts processing the commands.

II. Active – Active mode:

- 1) PM (SM) sends an ECHO frame to the SM (PM) is equivalent to Controller A of PM (SM) sends an ECHO frame to Controller A of SM (PM), and Controller B of PM (SM) sends an ECHO frame to Controller B of SM (PM).
- 2) If PM (SM) has all the resources to process a command, then SM (PM) sends an ECHO frame to the PM with the ECHO data indicating that processing command with OX-ID x_i , and the SM (PM) will yield and send an ACCEPT frame indicating that SM (PM) is dropping command with OX-ID x_i from its queue. This is an implicit AutoVolume Ownership Transfer. In the case of a race condition (that is both modules send the ECHO frame for a particular command), the PM has a higher priority in processing command than the SM. This means if both modules receive an ECHO frame for the same command, then the SM will yield and send the ACCEPT frame.

Contribution to the State of Art:

The method of this invention contributes the following aspects to the state of art:

- 1) Redundancy at the array controller module level which can recover from the following failures:
 - a. Controller module failure, which prevents both array controllers from accessing the drive trays. These failures can be related to power supplies failure, Fans failure, mid plane failures, and others.
 - b. Active controller fails in such a way, which prevents the alternate controller from accessing the drive trays.
 - c. Standby controller fails in such a way, which prevents the active array controller from accessing the drive trays.
- 2) An Automatic method to transfer the volumes ownership from the primary array controller module to the secondary array controller module or vice versa. Without any user intervention
- 3) Almost double the I/O throughput of an array system by using the bandwidth and the dual paths of the two redundant array controller modules.

Frequently Asked Questions:

- 1) How does this method handle caching?

Answer: It should be noted that if the array module is caching data then the two modules have to be in the active-active state. This way, if one of the controller caches a data for particular exchanges and it's the controller or the module fails, then the redundant controller on the second module will flushes its data to the drive bank from its cache because the original exchange was forwarded to the A controllers on both modules and the same for B controllers. Therefore, the redundant modules will handle this issue by the virtue of exchange broadcast and by the monitoring activities which happen between the multicast groups on both modules.

- 2) Does this method impact the performance?

Answer: Broadcasting the exchanges to the A or B controllers does not require any additional bandwidth from the host because the Multicast is built into different network protocols. Therefore, there will be very small overhead to lock step the As and Bs controllers on both modules.

- 3) How does the two controllers module present its information back to the host?

LSI LOGIC CORPORATION INVENTION DISCLOSURE FORM
(Mail to "Intellectual Property Law Dept. Attn: New Invention Disclosure" – Mail Stop AD-106)
Please attach additional pages where required

Answer: In the active – standby case, only the active presents its devices back to the host. In the case that module 1 fails, then standby module disables module 1 ports, becomes active, goes through a rest cycle, present its devices to the host, and then process the exchange.

In the active – active state, both modules presents its devices (duplicate devices) to the host adapters but because the As controllers have multicast addresses, then the host adapter presents the primary modules devices to the O/S. This aspect is built in the protocol.

In the case that module 1 fails, and then alternate module disables module 1 ports, becomes active, goes through a rest cycle, present its devices to the host, and then process the exchange.